



Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at <http://about.jstor.org/participate-jstor/individuals/early-journal-content>.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

SOME ACHIEVEMENTS IN THE ESTABLISHMENT OF A STANDARD FOR THE MEASUREMENT OF ENGLISH COMPOSITION IN THE BLOOMINGTON, INDIANA, SCHOOLS

EARL HUDELSON
Indiana University

Whoever experiments today with standards for measuring efficiency in composition is a pioneer. Standards for even such mensurable subjects as mathematics, spelling, and writing are yet young; while measurements for such a subject as composition, into which certain immensurable qualities enter, are literally new. Until yesterday, teachers of English believed that the only competent standard by which to measure composition was the Golden Rule. To some extent, indeed, those teachers of yesterday were right. There is at least one element in discourse which should be measured by no baser standard. Personality, originality, appreciation, individuality—call it what you will—that quality which stamps the product with the peculiarities of the producer—can never be scientifically measured; and any attempt to do so will only confuse the teacher and stifle the enthusiasm of the pupil. There are, however, fundamental qualities, such as spelling, punctuation, and rhetorical principles, which can be measured; and whereas we were formerly satisfied to rely upon our individual tastes for judging all qualities, we now are coming to believe that the more elements we apply to a definite standard and the fewer we leave to our several temperaments, the fairer will be our reckonings. It is the judging of these mensurable qualities for which we feel the need of a scale, and which we are today attempting to reduce to a scientific basis.

To see that such a scale is needed requires but a glance at such results as are pictured in Chart I.

The bent lines trace ten themes through the rankings of three teachers who did not use a scale. It will be seen that no one of

these compositions received a concurrent rating. Theme No. 604, which was ranked first by one reader, was placed tenth by another reader and fourth by a third.

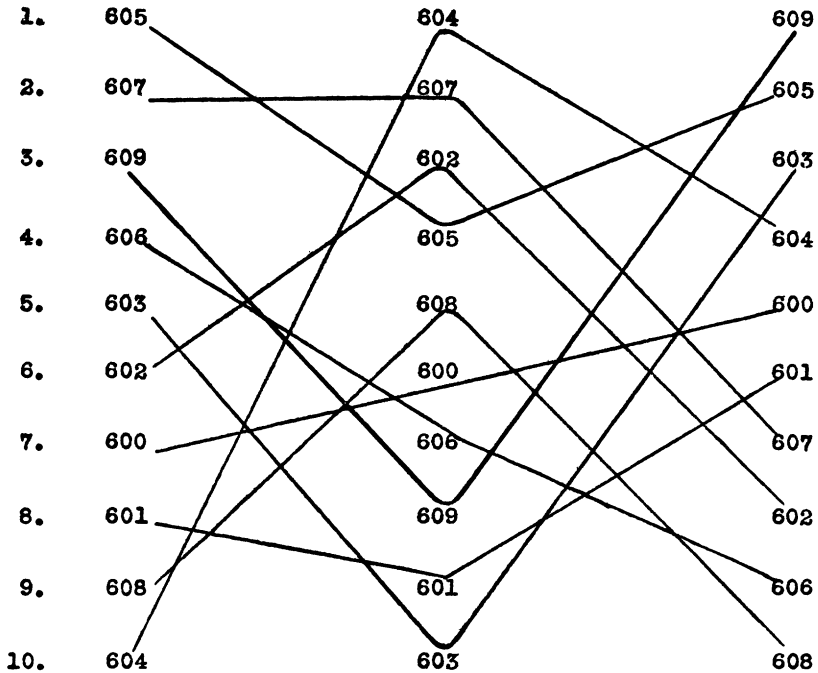


CHART I

Table I, which records the grades of eight compositions by ten graders each, shows that, while the average is not so bad, in one

TABLE I

706.....	55	88	71	80	60	90	86	85	90	70
709.....	70	89	78	90	83	94	85	85	92	70
713.....	40	84	85	76	50	95	82	70	75	65
726.....	80	94	75	92	80	88	91	90	95	90
740.....	75	87	79	90	70	70	81	79	85	75
761.....	40	92	72	86	75	92	80	80	60	75
762.....	25	75	50	42	30	45	40	40	30	35
794.....	30	75	62	38	40	45	45	35	45	40

case the same theme was ranked at 40 per cent by one grader and at 95 per cent by another. One grader passed all eight of the

pupils, while another passed only two of them. The pupil who was graded all the way from 40 per cent to 95 per cent was given, by the four teachers who failed him, an average of 56 per cent, while the other six graders allowed him an average grade of 83 per cent.

A moment's introspection will convince us that we do not grade compositions according to mathematical laws; that we do not give 100 per cent to the theme which is twice as "good" as the one to which we give 50 per cent; and that we do not visit a 30 per cent grade upon the pupil whose composition is one-third as good as the theme we grade 90 per cent.

I am not sure that we should immediately take upon ourselves the whole blame for, or to be surprised at, these chaotic conditions. Until we have evolved a standard by which to measure our conceptions of "good" and "bad," we can no more hope to reduce this chaos to order than to compute the distance from east to west. Were an Eskimo and a Brazilian to meet in Chicago, the former would complain of the heat while the latter would insist that he was freezing. Two English teachers meeting over the same set of compositions will, because of opposite former conditions, disagree upon the quality of the pupils' work. A teacher's ideal is a personal, individual one, and there are as many different ideals in composition as there are of heaven.

Realizing these chaotic conditions, Miss Kerr, principal of the departmental building, and I gave two tests to 800 pupils of the sixth, seventh, eighth, ninth, and tenth grades in the Bloomington, Indiana, schools. These tests were based upon the Harvard-Newton standard. The first one, held in February, 1915, was a test in description, and the report appeared in Vol. XIII, No. 11, of the *Indiana University Bulletin*. The second test, given in May of the same year, was in exposition. In both cases we substituted for the topics in the original scale, subjects as nearly equivalent as possible, but more suitable to our locality. We did not think it fair, for instance, to ask our pupils to describe a storm in a fishing village, when the angling of most of them had been confined to yearning for mud-cat in Beanblossom Creek. The topics offered for the two tests were:

DESCRIPTION

1. A Body of Water.
2. Some Person in Bloomington.
3. Grandmother.
4. An Old-Fashioned House.
5. A Picture.
6. A Public Building in Bloomington.
7. A Wreck.

EXPOSITION

1. Our Grading System.
2. How Stone Is Quarried.
3. How to Make ———.
4. How to Play ———.
5. Why ——— Is My Favorite Book.
6. Why ——— Is My Favorite Study.
7. How an Asphalt Street Is Made.
8. How Bloomington Should Dispose of Its Garbage.
9. How Bloomington Can Be Made a Really "Dry" Town.
10. How to Use a Dictionary.

Ten minutes were given for choosing and organizing the subjects, and thirty minutes were allowed for writing the compositions. Upon the first sheets were written only the subject, name, sex, age, grade, school, teacher, and date.

Regulation paper and ink were used, and no means of identification were allowed on any but the first pages, which were later removed from the compositions proper. These 800 themes were then jumbled into eight equal sets and each paper in each set was graded twice by each of three graders, no grader reading more than one set. The first reckoning was made independent of any scale; the second with reference to the Harvard-Newton standard. The grading was done by the English teachers of the departmental and high schools of Bloomington and by ex-teachers and prospective teachers in my university class in the teaching of high-school English.

In both tests the girls generally outdid the boys. The latter did their best work on such outdoor or violently emotional subjects as "A Body of Water" or "A Wreck," while the girls excelled in gentle sentiment about "Grandmother." In exposition the

boys were best in constructive themes, such as "How to Do _____," or "How to Make _____," or "How to Play _____," and left the purer reasoning of such topics as "Our Grading System," or "Why _____ Is My Favorite Study," or "Why _____ Is My Favorite Book," to the gentler sex. They were all responsive to the crying need of making Bloomington a really dry town.

The most interesting result for us, however, was our finding that while the graders were usually almost unanimous on the best and poorest compositions and at wide variance on the mediocre ones, yet in some cases the same theme was marked best, poorest, medium. As we have seen, one composition received grades of 40 per cent and 95 per cent without the use of a scale. This is pretty convincing evidence that something is wrong; and when the use of a scale reduces the divergence of this and hundreds of other compositions, it is another sign that a standard will do no harm, and will probably do good.

We found that the use of the Harvard-Newton scale reduced variance in a teacher's grades on the same theme from one time to another. Moodiness, or temperament, too often decides Johnnie's fate. If a teacher gets up on the wrong side of the bed, she may grade a composition 10 per cent lower than she would have had she been in a merrier mood. If Johnnie is near the danger-line, his chances depend upon his teacher's mood. While a standard can hardly be expected to eliminate temperament entirely, it will minimize the evils flowing therefrom.

Our experiments indicate that a scale will correct the injustice done to good students by teachers without the courage of their convictions. Without a standard, such teachers tend to inflate the grades of poor themes, to keep their writers within sight, at least, of the passing-mark. This works a hardship at all times on the pupils who really have good compositions; and when, because of a grouch, the teacher fails to augment the poor student's grade, that student feels, perhaps rightly, that an injustice has also been done him.

Table II, typical of most cases, reveals a tendency to telescope the lower end of the class up around the passing-mark. With the

use of a scale the percentage of poor grades is properly increased and a more normal curve denoted.

TABLE II
(Based upon ninety-five papers)

Percentage	With Scale	Without Scale	With Scale	Without Scale	With Scale	Without Scale
10- 20....	0	0	0	3.3	0	7.3
20- 30....	3.1	3.1	4.2	1.1	2.0	5.2
30- 40....	10.7	4.2	2.1	2.1	3.0	6.2
40- 50....	11.0	5.3	2.1	15.0	4.2	14.0
50- 60....	16.0	2.1	10.0	15.7	6.0	15.7
60- 70....	13.0	6.3	21.7	13.5	16.8	20.0
70- 80....	22.0	24.0	22.0	24.0	40.0	20.0
80- 90....	20.0	36.6	29.0	18.0	26.0	10.5
90-100....	4.2	18.0	8.9	7.3	2.0	1.1

By classes, the range of grades should, on the contrary, tend toward a reduction. After the standard for a class is once carefully established, the teacher's purpose should be to try to keep the majority of pupils around or above that standard. Then if a pupil is not within a reasonable distance above or below that ideal minimum, he is either too good or too poor for that class and should be placed in another.

This I take to be the happiest function of a standard of measurement. We must start somewhere, and rather than go through all that has been done to work out established scales, we can use such standards as the Hillegas or the Harvard-Newton to fix our units of measurement, much as the zero and boiling-points are established on a new thermometer by measuring it with an authentic one. In composition, by establishing one or two points we can, from them, derive the other degrees. These need not be fixed upon a percentage basis; in fact, my belief is that a standard should be made to run from 10 to 120 in point or degree of difficulty; and the ideal minimum for the first grade would then be 10, for the second 20, for the third 30, and so on, up to 120 for the twelfth year. With this as a basis, our problem would then become one of choosing typical models for each year. These models should be made accessible to the pupils, that they may learn just what is considered a fair theme for them. Then, I repeat, the promise of promotion will induce them to excel the model.

One shortcoming of the Harvard-Newton standard is its lack of variety of models. Practically all of the average pupil's writing after he is through school will be in epistolary form; and no one sticks to a single discourse. Yet the Harvard-Newton scale contains no samples of letters and no intentional types of mixed discourse.

The minimum grade in the Harvard-Newton test in description is about 44 per cent and in exposition about 39 per cent. Before this standard should be used permanently, or another established from it, it should be completed downward. Otherwise conscientious teachers who undertake to grade below the minimum of the scale will feel their supports knocked out from under them, as it were, and will tend to keep their grades within purview of the scale.

No standard, probably, will ever be made that will be equally suited to all schools. Just as we found it advisable to substitute topics in our tests at Bloomington, so I believe each school should have essentially its own standard, in order that compositions typical of the locality may be used in the scale.

A common tendency in work of this kind is to discontinue the experiment just when results are imminent. We then grumble, like the stingy farmer who gradually decreased his horse's rations in hopes that he might finally eliminate this expense. The farmer, in reporting the project, said, "I got along beautifully till I got to the place where Dobbin didn't need any more; then the old fool died!"

No one can reasonably say he is familiar with the scale until he has used it assiduously for months; and experiments based upon any shorter acquaintance with the standard, however sincere, cannot be called strictly reliable.

"But," the average teacher will ask, "how, with English club, high-school paper, annual, literary societies, debates, plays, compositions, notebooks, library work, outside reading, tests, examinations, and a few classes a day—to say nothing of three bolted meals and a few hours of sleep—can I find time to pursue these investigations, however badly they are needed?" It is a problem. The grading of papers in such tests is the least of the problem, too, because teachers are about the only ones who are capable of rating

compositions reliably. It is not necessary, however, that teachers do the compiling of statistics; indeed, it is better for them not to, for there are others—efficiency experts—who are trained to do such work quickly and accurately, and who are needed in every city, the size of Bloomington and larger, to put the results of any tests into a form that can readily be used by anyone who desires the results.

Our investigations have convinced us that a standard of measurement is needed in composition; that each school should have, in the main, its own scale; that this scale should be generous but typical; and that it will then organize the grading of compositions scientifically, minimizing unequalization and variability, and giving Johnnie an equal chance, under whatever teacher he may happen to be.